Lexical gaps in the vocabulary structure of a language

Rajendran Sankaravelayuthan
Amrita University
Coimbatore
rajushush@gmail.com

1. Introduction

Linguists consider word as a crucial unit in their description of language. While doing so they mostly focus on those words that are recognized as part of the vocabulary of a language. Sometimes it is relevant to consider the words that are not part of the vocabulary. They can be referred as non-existing words. In lexical semantics it is customary to talk about lexical gaps instead of referring to non-existing words. The non-existing words are indications of "gaps" or "holes" in the lexicon of the language that could be filled.

Lexical gaps are also known as lexical lacunae. The vocabulary of all the languages including English and Tamil shows lexical gaps. For example, English noun *horse* as a hypernym, incorporates its denotation both *stallion* (male horse) and *mare* (female horse). However there is no such hypernym in the case of *cows* and *bulls*, which subsumes both *cow* and *bull* in denotation. The absence of such a hypernym is called a lexical gap. Lyons (1977, pp. 301-305) addresses lexical gaps from a structuralist's perspective. A structuralist defines lexical gaps as slots in a patterning. Wang (1989) defines lexical gaps as empty linguistic symbols and Fan (1989) defines them as empty spaces in a lexeme cluster. Rajendran (2001) defines lexical gap as a vacuum in the vocabulary structure of a language.

We always encounter the lexical gaps when we try to translate one language into another or develop a bilingual or multilingual dictionary or lexical data base like wordNet or thesaurus or ontology for the vocabulary of a language.

2. Lexical idiosyncrasies

One will come across various types of idiosyncrasies or discrepancies if one makes a contrastive analysis of a source and a target language. Bentivogli & Pianta (2009) gives the following as a summary of the most common idiosyncrasies:

2.1. Syntactic divergences

The syntactic discrepancy arises when the translation equivalent (TE) does not have the same syntactic ordering properties of the source language word.

e.g.

- English is shows SVO word order whereas Tamil shows SOV word order.
- English is prepositional language whereas Tamil is a postpositional language.
- Complementation is English is different form complementation in Tamil.
- Relative clause in English is different form relative clause inTamil

- The phrase *king of England* in English needs to be translated as *ingkilaant-in arasan* ('England-possessive case marker king') in Tamil.
-

The list will extend to a considerable extent.

2.2. Lexicalization differences

Lexicalization differences come to fore when the source and target languages lexicalize the same concept with a different kind of lexical units (word, compound or collocation) or one of the two languages has no lexicalization for the concept (lexical unit vs. free combination of words). The latter case is called lexical gap. Take for example *bicycle*; bicycle has been introduced to Tamil culture from outside. Tamil has barrowed the word *caikiL* along with the vehicle, *cycle*. Later on Tamil tried to coin its own indigenous name from its own units of meaning. Thus many names are coined for *bicycle*: *miti vaNTi* which literally means 'vehicle which need to be peddled', *untu vaNTi* 'vehicle which need to be pushed', etc. Similarly *car* is taken with its foreign name *kaar* 'car'. Later on *ciRRuntu* (which literally means 'small vehicle) is coined. *Bus* is taken into Tamil culture with its foreign name *pas* 'bus'. Later *peerundtu* is coined (which literally means 'big vehicle'. But Tamil shows vacuum or gaps in representing the certain parts of these vehicles. Kinship terms have full of these examples. For the English kinship *uncle*, there are many equivalents in Tamil each denoting different kinship concepts. So if you go from Tamil to English, you will realize that there are many lexical gaps in English.

2.3. Divergences in connotation

The TE fails to reproduce all the nuances expressed by the source language word. For example *knowledge* in English cannot express all the nuances expressed by *aRivu* in Tamil. *Philanthropy* cannot express all the nuances expressed by the Tamil word *tarmam.* Similarly *paavam* and *puNNiyam* in Tamil cannot be equated respectively with 'sin' and 'blessing' in English

2.4. Denotation differences

The denotation difference appears when the denotation of the source language word only partially overlaps the denotation of the TE. For example, English *finger* only partially overlaps with Tamil as *finger* as it denotes only the terminal part of the hand and not the terminal part of a leg; English makes use of *toe* to denote the terminal part of the leg. Tamil makes of *viral* to denote both the terminal part of the hand and the terminal a part of the leg.

Only the first two idiosyncrasies are relevant to translation as they imply lack of cross linguistic synonymy. They are represented as:

• Lexical gaps: Lexical gap denotes an instance where a language expresses a concept as a lexical unit or word while the other language expresses it with a free combination of words. For example, the word *borrower* is referred in Tamil by the phrase *kaTan vaangkupavar* 'one who gets loan'.

• Denotation differences: Denotation difference arises when the TE of a source language exists but it is more general (generalization) or more specific (specification). In the former case the TE is a sort of cross-

linguistic hypernym of the source language word (ex. Tamil *viral* = English *finger* or *toe*) and in the latter case it is a cross linguistic hyponym (for example, English word *lion*, which functions as a general term for lion and male of lion is equaled by *cingkam* in Tamil which is only a general term for *lion*; 'male lion' is denoted by the phrase *aaN cingkam*)

3. Defining a lexical gap

A competing term for 'lexical gaps' is ' lexical holes'. The two terms are alternatively used in the literature available on the topic. However, 'lexical gap' as a term is widely used than 'lexical hole'. The definition of lexical depends upon whether we talk about lexical gap within a language or across the languages. As far as translation is concerned, the lexical gaps across language are crucial ones. Of course the lexical gaps within the language too help us to understand the lexical gaps across the languages in clear terms.

There is a unanimous agreement between linguists and translation specialists of what a lexical gap means. Trask (1993:157) defines lexical gap as "the absence of a hypothetical word which would seem to fit naturally into the pattern exhibited by existing words". The pioneer in field semantics, Lehrer (1974:95) states that the term 'lexical gap' is multiply ambiguous as it has been applied to all sorts of instances where a word, in one way or another, is missing. A lexical gap means the absence of lexicalization of a certain concept. A concept is lexicalized when a language has a lexical item to express the concept. The lexical item could be a single word, a complex word, an idiom or a collocation. The existence of a lexical gap will be noted only when a concept lack lexicalization and is expressed by a free word combination or any other transformation (e.g. omission, translation by different parts of speech, etc.). Thus the multiword expression X is not a lexical gap, because it is a fixed expression in a language, while Y is a lexical gap, because it is a free-word combination.

In the case of lexical gap across languages, lexical gaps are considered as instances of lack of lexicalization identified in a language while comparing two languages or in a target language during translation. The problem seems to be minor and clear. But, after going through the linguistic literature on lexical gap one gets rather an opposite impression. The problem is on lexicalization which is explained from linguistic perspective. The definition of lexical gaps is based on linguists' practical requirement as well as their understanding of the process of lexicalization. Thus, lexical gaps are realized across the languages in the following three instances: one is when a source-language word does not have a direct equivalent without going into details about the notion of a direct equivalent itself (Janssen 2004); the second is when a source-language word rendered by a target-language word is rendered by a target-language phrase without distinguishing it from idioms and collocations (Arnold et al 1994, Santos 1993); and the third is when a concept is not encoded by a lexical item, i.e. by a word, a complex word, an idiom and a collocation (Bentivogli and Pianta 2000, Bentivogli et al 2000). The main difference between these definitions can easily be noticed: the specificity of a lexical item.

A lexical gap is identified on the level of one meaning. It is not identified on the level of a lexeme, which is usually polysemous. In translation we deal with one meaning with reference to context specific only to a particular meaning. A translator is interested in individual meanings. He is not bothered about the semantic structure of a word. Therefore the lexical gaps are identified on the level of individual meaning. In general we can say that a lexical gap is a concept which is not lexicalized by a lexical item (single word, complex word, an idiom and collocation) in a language due to cultural, language/systematic

or other reasons. Lexical gap is identified while comparing or translating, individual meanings of lexical items in two languages. It is usually expressed by a free word combination in translation. One of the reasons to study lexical gap is that it is difficult to identify them in advance. Only during translation one understands that the target-language lacks a certain word. A dictionary in such cases provides a mere explanation of the concept encoded by a source language. Unfortunately, such meaning explanations usually are not good in natural language use.

4. Typology of lexical gaps

The study of lexical gaps starts with the work by Chomsky (1965) and Chomsky & Halle (1965). They distinguish between accidental gaps and systematic gaps. Accidental gaps are words that do not exist but could be reasonably expected to exist; on the other hand systematic gaps are words that are not even expected to exist since they violate the rules of what a "good" word is. However, the term lexical gap is reserved only for the accidental gaps in much of the subsequent works.

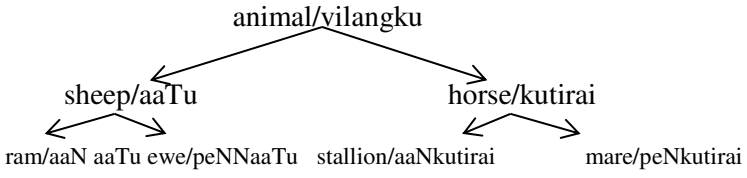
The accidental gaps in the work of Chomsky and Halle are segments or strings of letters that could possibly form words. Such gaps are called formal gaps, sometimes also referred to as morpheme gaps. DAY TRANSLATIONS (2018) opines that lexical gaps occur in several types. They are realized at phonological (e.g.*pkly/pkli/), morphological (e.g.*ungood), syntactic (e.g.*informations) or semantic (e.g. *male dog) levels. A significant part of the more recent work on lexical gaps, however, deals with semantic gaps. A semantic gap is, in the words of Lehrer (1974), "the lack of a convenient word to express what (the speaker) wants to speak about.", although also words that are possible but not (yet) convenient are considered semantic gaps. A semantic gap is a notion for which there is no word, whereas formal gaps are "words" that do not refer (to any notion). As with formal gaps, we can in principle distinguish between semantic gaps that are accidental, and semantic notion for which no word can exist because they violate the rules of what a "good" notion (for lexicalization) is.

The coarse taxonomy of non-existing words given by Janssen (2004) is given below:

	Accidental	**Systematic**
Formal	Formal gap	Impossible lexical entry
Semantic	Semantic gap	Non-lexicalizable notion

An overview of types of lexical gaps given by Janssen (2004) is shown below:

Morpheme gap	A sequence of segments that is permitted by phonological rules but not found. Fillers: possible words. e.g.*pkly/pkli.
Morphological gap	A word that can be generated from an existing word by productive morphological rules. Mostly understood as derivational rules, and therefore also called derivational gaps. Fillers: potential words. e.g.**ungood*

Paradigm gap	A morphological gap in the inflectional morphology. e.g. *buy-bought- bought* (against *take-took-taken*) Paradigm gaps across languages are of great concern for translation. • The absence of subject agreement markers in finite verbal forms of English when compared to Tamil. e.g. *vant-een* 'came-I, *vant-aay* 'came-you', *vant-aan* 'came-he', *vant-aaL* 'came-she', *vant-aar* 'came-he' • The absence of three types of degree words in Tamil against English: e.g. good – better - best. • The absence of agreement marked adjectives against unmarked adjectives in English. E.g. *nalla-van* 'good-he', *nalla-vaL* 'good-she', *nallav-ar* 'good-he
Semantic/functional gap	A lack of a word to express what a speaker might want to talk about.**male dog*
Taxonomic gap	A gap in the taxonomic structure. Fillers: pseudo-words
Translational gap	A word in one language for which no lexical unit exists in another that expresses that same meaning. Fillers: untranslatable words. E.g. absence of a hypernym subsuming verbs of motion in land. Another example: 

5. Lexical gaps and Semantic fields

The principles of semantic field contribute to the proper understanding of lexical gaps. The proponents of the semantic field theory (Lehrer 1974a, 1974b; Lyons 1977) declare empty spaces in a structure which is also related to absence of lexicalization as the essential feature of a lexical gap. In this approach different fields: taxonomies, hierarchies, clusters, grids, linear structures and matrixes help to organize the lexicon into conceptual structures where the missing structural part is then best observable and studied with relation to the other units in the field. Other approaches, Bentivogli and Pianta for example, favour contrastive lexicological studies where a lexical gap is identified as a missing translational equivalent in a target language to a lexical item in a source language. The study of lexical gaps has received increased attention recently, because of the present-day demand for the translation of all kinds of documents, statutes, provisions, regulations, licences, contracts and others.

In the light semantic field or lexical field we can investigate clearly about the lexical vacuums or lexical gaps. The theory of semantic fields emerged heavily influenced by de Saussure's structuralism and German idealism. Trier, who is a pioneer in the lexical field analysis, opines that lexical fields are neatly structured. The whole vocabulary is organized in fields. He introduces the notion of concepts and fields and conceptual fields and advocates that when concepts change in our heads, meaning of a lexeme also changes. Semantic field demonstrate vocabulary organization on the paradigmatic level. The basic assumption is that the vocabulary of a lexical field is an integrated system of lexemes which are interrelated in meaning. The whole of lexical field consists of a large number of semantic fields which accumulate lexemes which are close in meaning. The vocabulary of a lexical field is a mosaic without gaps or overlaps. His followers disregarded conceptual field. They preserved very neat and rigid structures, but couldn't explain how and why lexical fields change.

It is often common for lexical gaps to come to the fore within semantic fields where there is a hole in the pattern i.e. "the absence of a lexeme at a particular place in the structure of a semantic field" (Lyons, 1977:301). For instance, the semantic field of temperature in English, as introduced by Conner (1983:43), consists of four terms: cold, cool; warm, hot. In some contexts these terms turn to be synonyms (e.g. cold/ cool water) and in other they are antonyms (e.g. cold/ *cool outer space). On the other hand, in Tamil the semantic field of temperature involves different terms where a lexical gap is easy to recognize.

The semantic field of temperature:

English: freezing, cold, cool, lukewarm, warm, hot, scorching.

Tamil : *kaTungkuLir*, *kuLir/taNu*, *vetuvetuppu*, *mitamaana cuuTu*, *nakaccuuTu*, *cuuTu*,

English: Fantastic, excellent, good, bad, awful, abysmal

Tamil: *mikananRu, nanRu, moocam, mikamoocam*

Semantic filed of young ones:

English: dog: puppy :: cat : kitten,

Tamil : *naay* : *kuTTi*::*puunai* : *kuTTi*

Semantic field of stages of butterfly:

English: egg: larva: pupa: butterfly

Tamil: *muTTai: laarvaa: piyuupaa: vNNattuppuucci*

Semantic field of colour:

English: Violet: indigo, blue: green: yellow: orange: red

Tamil: *vailaTTu, iNTIkoo, niilam, paccai: manjcaL: aaranjcu, civappu*,

The lexical framework of any language is often built in terms of semantic fields (e.g. kinship relations, colour terms, military ranks etc.), sense relations (e.g. hyponymy, synonymy, antonymy etc), collocation, idioms and relational opposites. The basic principle behind the availability of certain lexis in a given language is its users' need. So, a lexical item referring to a particular object or concept can be found in one language, but it is absent in another. Bentivogli and Pianta (2000) are of the opinion that a lexical gap occurs whenever a language expresses a concept with a lexical unit whereas another language expresses the same concept with a free combination of words. Lyons (1977:303) maintains that lexical gaps are attributed to unlexicalized concepts or objects across languages. For instance, the distinction between dead humans and dead animals lead to the coinage of two lexical items referring to them as 'corpse' and 'carcass' respectively as a consequence of institutionalization. However, there is no word referring to dead plants.

Lexical gap ("hole in the pattern") indicates the absence of a particular lexeme in a point in a particular lexical field. According to Trier there are no gaps in the system. If they arise (by conceptual innovation), they are quickly filled by borrowing or by extending the meaning of an existing lexeme. It should be remembered in the context that according to Chomsky there are no gaps in the system. Gaps appear when you compare languages. Languages show cultural gaps. As one language is culturally different from another language, it is likely that the cultural items of one language may not be found in the other language. So it is needless to say that lexical gaps are inevitable in the vocabulary structure of a language. The main reason for lexical gaps is the absence of lexicalization which is not easily pinned down. However, a major group of lexical gaps can be explained by social and cultural differences of source and target language users. A lexical gap in a target language is identified when its users cannot know the concept encoded by a source language.

Lexical framework of any language is often built in terms of semantic fields (e.g. kinship relations, colour terms, military ranks etc.), sense relations (e.g. hyponymy, synonymy, antonymy etc), collocation, idioms and relational opposites. The basic principle behind the availability of certain lexis in a given language is its users' need. So, it is possible to find a lexical item referring to a particular object or concept in one language, but it is absent in another. Bentivogli and Pianta (2000) opine that a lexical gap occurs whenever a language expresses a concept with a lexical unit whereas another language expresses the same concept with a free combination of words. Lyons (1977:303) maintains that lexical gaps are attributed to unlexicalized concepts or objects across languages. For instance, due to the cultural institutionalization of the distinction between dead humans and dead animals, two lexical items are coined referring to both as 'corpse' and 'carcass', respectively. However, there is no word referring to dead plants.

Rajendran who prepared a thesaurus for Tamil (Rajendran, 2001) based on Nida (Nida, 1975) who developed a thesaurus dictionary for bible translation on the principles of componential analysis makes a detailed study on lexical gaps or vacuums in the vocabulary structure of Tamil (Rajendran, 2000) .

6. Semantic Structure of Vocabulary

'Semantic structure of Tamil vocabulary' (1998) was worked out by me as my postdoctoral research work (Rajendran, 1982) by making use of the classification of vocabulary advocated by NIda (1975). Nida's classification is based on the principles of componential analysis propounded by him. Nida's (1975) work culminated into a thesaurus for Greek language for bible translation. I have followed

the same schema when I converted 'Semantic structure of Tamil vocabulary' into Tamil Thesaurus (Rajendran, 2001). While attempting semantic structure of Tamil vocabulary. I realized that there are huge vacuums or lexical gaps in Tamil vocabulary. Many lexical domains have no vocabulary less vocabulary in Tamil. Say for example many domains of knowledge such as science, astronomy, law, etc do not have sufficient vocabulary in Tamil to portray them in a taxonomic fashion. While attempting to classify Tamil vocabulary based on lexical relations such as hyponymy and superordinate relation and part and whole relation I could not find superordinate terms for a number sets of lexical item which I feel like classifying under superordinate terms. Also while comparing lexical items found in Tamil with lexical item in English, I come across many vacuums as well as many mismatches between them. I would like to call these vacuums as well as the mismatches as lexical gaps. I would like to discuss in this section the lexical vacuum or lexical gaps. The study of lexical gaps can be approached from the point of view of lexical fields, as suggested by Lehrer (1974) and Cruse (1986) among others.

Nida (1975a) who was concerned with the preparation of a thesaurus dictionary for Greek gives the following as the tentative hierarchical classification of the lexical items (Nida:178-186).

I. Entities

A. Inanimate

1. Natural

a. Geographical

b. Natural substances

c. Flora and plant products

2. Manufactured or constructed entities

a. Artifacts (non constructions)

b. Processed substances: foods, medicines, and perfumes

c. Constructions

B. Animate entities

1. Animals, birds, insects

2. Human beings

3. Supernatural power or beings

II. Events

A. Physical, B. Physiological, C. Sensory, D. Emotive, E. Intellection, G. Communication, G. Association, H. Control, I. Movement, J. Impact, K. Transfer, L. Complex activities, involving a series of movements or actions

III. Abstracts

A. Time, B. Distance, C. Volume, D. Velocity, E. Temperature, F. Color, G. Number, H. Status, I. Religious character, J. Attractiveness, K. Age, L. Truth-falsehood, M. Good-bad, N. Capacity, O. State of health, etc.

IV. Relationals

A. Spatial, B. Temporal, C. Deictic, D. Logical, etc.

The major classification of vocabulary by Nida (1975) is to group it into four: entities, events, abstracts and relationals. He further classifies each domain into subgroups based on componential semantic features. His classification is based on referential meanings and it is not possible to obtain one to one correspondence between the semantic domain of classes and the grammatical classes. There are, of course certain parallel between them, "since on some level of the deep structure entries tend to be represented by nouns, events by verbs, abstracts by qualifiers, and relations by a number of different features: particles, affixes of case, word order, etc." (Nida, 1975.a:176). There is evidently a clear logic governing such a classification, though it is not difficult to construct an alternative scheme, whilst retaining a similar line of thought with reference to the vocabulary of a language. Nida's universal semantic classification can be adapted for Tamil without much drastic changes, though one may come across a number of problems while doing so.

7. Lexical gaps in vocabulary structure

As mentioned in the introduction, while trying to classify vocabulary based on the schema propounded by Nida (1975), I felt Tamil vocabulary has many vacuums or lexical gaps in its vocabulary structure. Some of the vacuums or lexical gaps are discussed below.

7.1. Lexical gaps in the domain of entities

7.1.1. Lexical gap in the domain of inanimate entities

7.1.1.1. Lexical gap in geographical domain

Geography is the branch of science which deals with celestial objects, space, and the physical universe as a whole. Tamil vocabulary shows many lexical gaps in celestial-atmospheric sub domain. Supernatural sub domain of Tamil is totally different form that of English. The concept of *navakirakam* 'nine planet' is quite indigenous which is different from that of universal solar system represented in English. The *navakirakam* consists of *cuuriyan* 'sun', *cantiran* 'moon', *cevvay* 'Mars', *putan* 'mercury', *viyaazan* 'Jupiter', *cukkiran* 'venus', *cani 'saturn'*, *raaku* 'a mythological planet', *keetu* 'another mythological planet'. In this system earth is not considered as one of the 'nine planets'; *cantiran* 'moon' which is a satellite of earth is considered as one of the nine planets; *curriyan* 'sun' which is a star is also considered as one of the 'nine planets'. According to this system these 'nine planets' move around earth. The regular solar system can also be represented in Tamil.

The sub domain 'politically defined areas' in Tamil too shows disparity with that of English.

7.1.1.2. Lexical gap in flora domain

As the flora differs from region to region, the said domain inherits a lot of lexical vacuums or gaps by default. There is an incredible number of different plants in the world. Some are familiar to Tamil soil or Tamil society. A sharp classification based on generic and specific features is not available for Tamil. Botanists have tried to separate plants according to selected traits: plants having seeds or vascular tissue, mono-cotyledon plants vs. dicotyledonous plants, flowering plants or non-flowering plants, fruit bearing plants vs. non-fruit bearing plants, plants with tap roots vs. plants with secondary roots, branching plants vs. non-branching plants, and so on. Plants have been grouped into twelve different phyla depending on these characteristics. Incidentally, learning about the types of plants also takes us on an evolutionary journey as plants emerged from aquatic systems and increased in complexity. Due to lack of information such classification is not attempted for the domain of plants in Tamil. Such sharp classification of plants in Tamil soil is yet to be attempted even in Botany text books in Tamil. In Tamil a good number of plants are named after their use in medicine.

I have made use of Madiyan and Chitraputran's (1986) book on *marainappeyar tokuti* (Volume on plants) for the vocabulary for flora domain. I have made use of the classification available in the book: trees are classified as big trees, long trees, medium trees, small trees; plants have been classified as plants, shrubs, big and long shrubs, smalls shrubs; there are other groups such as creepers and greens, Though they have given botanical names of the flora, I could not attempt a full-fledged scientific classification based on Aristotelian principle of genera and species from the information available in the volume. Madavan and Chitran (1986) have concentrated on the collection of synonyms rather than generic-specific classification. There are many gaps or vacuums to attempt such classificatory system.

7.1.1.3. Lexical gap in the domain of manufactured or constructed entities

Many lexical gaps can be found in the domain of manufactured or constructed entities too. For example, there are many indigenous weapons and tools belonging to Tamil society which do not have equivalents in English. The reverse may be also true.

7.1.1.3.1. Lexical gap in the domain of vehicles

Many vehicles are introduced to Tamil society from outside. Tamil either coin indigenous words to denote them or adopts names of the foreign vehicles along with them. Take for example bicycle; bicycle has been introduced to Tamil culture from outside. Tamil has barrowed the word *caikiL* along with the vehicle *cycle*. Later on it tried to coin its own indigenous name from its own units of meaning. Thus many names are coined for bicycle: *miti vaNTi* which literally means 'vehicle which need to be peddled', *untu vaNTi* 'vehicle which need to be pushed', etc. Similarly *car* is taken with its foreign name *kaar* 'car'. Later on *ciRRuntu* is coined. *Bus* is taken into Tamil culture with its foreign name *pas* 'bus'. Later *peerundtu* is coined. But Tamil shows vacuum or gaps in representing the part of these vehicles. Parts of car, bus, train, aero plane and ship are not amiable to Tamil.

7.1.1.3.2. Lexical gap in the sub domain of tools and instruments

Many tools and instruments of indigenous origin cannot be glossed by English equivalents. There are many varieties of cutting and digging instruments as well as tools for carpentry and sculpture of Tamil origin which do not have exact English equivalents. The reverse can be true too. It is difficult to find

standards in these domains. If you consider the vocabulary of English as standards, one can say that there are lexical gaps. But if one takes Tamil as standard, you will find lexical gaps in English.

7.1.1.3.2.3. Lexical gap in the sub domain of clothes

Clothes and dresses differ from region to region. There are many clothes and dresses which may not have glosses of equivalents in English and reverse is also true. As said in the previous case, here also it is difficult to find standards in this domain. If you consider the vocabulary of English as standards, one can say that there are lexical gaps. But if one takes Tamil as standard, you will find lexical gaps in English.

7.1.1.3.2.4. Lexical gap in the sub domain of processed substances

The processed substances like foods, medicines and perfumes too differ from region to region. Finding universals in this domain is a difficult task. The Tamil food item *toosai* ('a food item made from the liquidized dough of rice and black gram') is *toosai* only and similarly *cake* is *cake* only.

7.1.1.3.2.5. Lexical gap in the sub domain of constructions

Constructions too differ from region to region. . Lexical gaps are expected in the domain of constructions too. The construction of temples found in Tamil culture cannot be matched with construction of churches in English society. The parts of a temple is quite unique to Tamil culture or Indian culture and may find their names in other culture. The indigenous concept *muRRam* 'similar to courtyard' and *tiNNai* 'a part in front of a Tamil house' do not have apt equivalents in English. The different types of houses in Tamil culture do not find their equivalents in English. Similarly construction of different types foreign to Tamil culture may not find their place in Tamil vocabulary system.

7.1.2. Lexical gap in the domain of animate entities

7.1.2.1. Lexical gap in the domain of animals, birds and insects

The vocabulary of animals, birds and insects found in Tamil represents Tamil region or Indian region. The animal, birds and insects found elsewhere do not find their place in Tamil vocabulary. Even then I have attempted Aristotelian way of classification by general and specific features. I tried to fill the gaps making use of information available in Wikipedia. Wikipedia gives, for example, different types animals which include different types of elephants, tigers, lions, snakes, etc., and birds and insects. Tamil vocabulary system will show huge vacuums or gaps in the domain of animals, birds and insects.

7.1.2.2. Lexical gap in the domain of Kinship terms

Kinship terms are culturally bound lexical items. So many mismatches or gaps are found in the domain of Kinship terms while glossing Tamil Kinship terms with English glosses or vice versa.

- Even in the basic concepts 'mother' and 'father' are very much mismatched between Tamil and English. *appaa* 'father' and *ammaa* 'mother' not only denote kinship relations they have meaning extensions which do not mach with the extension *father* and *mother* in English.
- *appaa* in Tamil can be combined with *peria* 'big' and *ciRiya* 'small' form *periyappaa*, ciRRappa denoting kinship terms which can be glossed as 'male person elder to father' and 'male person

younger to father'. Similarly *periyammaa* denotes a kinship concept which can be glossed as 'female person elder to mother' and *cinnammaa* denotes kinship concept which can be glossed as 'female person younger to mother'

- For English *word* uncle there could be many kinship concepts in Tamil. For example *uncle* could be *maamaa* 'mother's brother', *periyappaa* elder brother of father' , ciRRappaa 'younger brother of father' or it can denote any elderly male person.
- Similarly for English *aunt* can include many kinship concepts in Tamil such as *athai* 'mother's sister', *periyammaa* 'elder sister of father' , *ciRRappaa* 'younger sister of father' or it can denote any elderly male person.
- Tamil *aNNan* has to be glossed as 'elder brother' and *tampi* has to be glossed as 'younger brother'; similarly *akkaa* has to be glossed as 'elder sister' and *tangkai* to be glossed as 'younger sister'

7.1.2.3. Lexical gap in the domain of Personal pronouns

Languages differ in the lexicalization of personal pronouns. You will find pertinent lexical gap if you mach the pronouns of Tamil with English. The following table will illustrate this:

Tamil	**English gloss**
naan	I
naam	we (including speaker)
naangkaL	we (excluding spkeaker)
nii	you
niir	you with minimum respect
niingkaL	you plural/ you with respect'
avan	'he'
avaL	'she'
avar	'they'
avarkaL	'they with respect'

For English *we* there could be two possible equivalents in Tamil. Similarly for English *you* there could be three possible equivalents in Tamil, the selection of which depends on the social status of the speaker and addresser.

7.1.2.4. Lexical gap in body parts

The domain of body parts too shows disparity between languages. Tamil and English mismatches in the delineation of body parts. The following table shows the matching of the major body parts:

Tamil	English equivalent or gloss
talai	head
kazuttu	neck
maarpu	chest
vayaRu	stomach

The concepts 'leg' and 'arm' are realized differently in Tamil and English. English makes a distinction between limp and leg. *Limb* is defined as 'an arm or leg of a person or four-legged animal, or a bird's wing'. *Arm* in English is defined as 'each of the two upper limbs of the human body from the shoulder to the hand'. *Hand* is defined as the terminal, prehensile part of the upper limb in humans and other primates, consisting of the wrist, metacarpal area, fingers, and thumb. Tamil makes use of two words: *kai* denotes the portion which includes *arm* and *hand* of English *kai* can be glossed as 'fore limb/upper limb' and *kaal* 'upper limb of human being or leg of animals' arm and leg. By definition of *hand* of English is not equivalent of Tamil *kai*. *Leg* can be equated with *kaal* of Tamil. The portioning of *arm* and *leg* also shows mismatches when we compare Tamil with English. In Tamil *kai* 'arm' is partitioned as *meeRkai* 'upper arm', *muuTTu* 'knee', *munkai* 'lower arm', *maNikkaTTu* 'wrist' *uLLangkai* 'palm' and *viralkaL* 'fingers'. English has the following as parts of *arm*: from shoulder to wrist is *arm* which can be divided into *upper arm* and *forearm;* the next portion is *wrist* followed by *palm*. If we compare English with Tamil, one can infer that there is no equivalent word in Tamil for denoting *arm* and *hand* of English. *Leg* is defined in English as 'the lower limb of a human being or animal that extends from the groin to the ankle'. As per this definition *leg* excludes the portion *foot* which consists of ankle, heel and toes. Leg consists of *thigh, knee, shin* and *calf*. In Tamil *kaal* is equivalent to that part which includes *leg* and *foot* of English. In English lower limb of human being which is equivalent of *kaal* in Tamil is portioned into *leg* and foot; *leg* in turn is portioned into *thigh, shank* and *calf*. *Shank* is defined as 'the part of the human leg between the knee and the ankle'. Shank is portioned into *shin* and *calf*. Shin is defined as 'the front part of the human leg between the knee and the ankle'; *calf* is defined as 'the muscular back part of the shank'. In Tamil *kaal* consists of *toTai* 'thigh', *muTTu* 'knee' *muunangkaal* 'shank', *kaNukkaal* 'ankle' and *paatam* 'foot' which in turn has *kutikaal* 'heel' and *viralkaL* 'toes'; Tamil does not make the distinction between finger and toe. Tamil equivalent of *shank* is *muzhangkaal,* of *shin* is *muzhantaaL*, ankle is *kaNukkaal* and *calf* is *'keNTaikkaal'* .

7.2. Lexical gaps in the domain of events

The lexical items organized under events too show many varieties of lexical gaps if we compare it with English vocabulary. A look at the English Thesaurus of English Dictionary will stands to substantiate this fact.

7.2.1. Lexical gaps in the sub domain of movements

While structuring verbs of movement Tamil (Rajendran 1978, 1983) I have found that the Tamil vocabulary is not as rich as English. For example while Tamil has only *naTa* 'walk' to denote the movement of walking, English has many: *walk, stroll, saunter, amble, plod, trudge, hike, tramp, trek, march, stride*, and *step out*. Similarly while Tamil has only *ooTu* 'ran' to denote the movement of running,

English has many: *race, rush, hasten, hurry, dash, sprint, bolt, dart, gallop, career along, tear along, charge along, speed along, jog along, scurry, scamper, scramble*.

7.2.2. Lexical gaps in verbs of cooking

While working on the semantics of cooking verbs (Rajendran 1976) in line with Lehrar's Semantic cosine, I have found out mismatches between Tamil and English. As the food items preferred by Tamils differ from that of English food items, the cooking process in the two languages vary.

Similarly verbs of seeing (Rajendran 1982), verbs of eating (Rajendran 1995), verbs of communication, etc too show lexical gaps while comparing the concerned Tamil vocabulary with English vocabulary.

7.2.3. Lexical gaps in the meaning extension or polysemy

Even if a verb from Tamil is considered as equivalent to one in English, one cannot say that all the senses denoted by the Tamil and English verbs are same. There could be sense gaps between the parallel lexical items of the languages. For example the meaning denoted by *naTa* 'walk' is extended to denotes 'go on' as given in the following example: *kaTai naRaaka naTakkiRatu* 'the (business of the) shop is going on well'. English *walk* does not express this sense. English *run* is used as shown in the following example: *he is running temperature* (i.e. *he is having fever*). *ooTu* 'run' does not show this extension of meaning. Similarly the extension found in Tamil *ooTu* 'run' may not find its place in its equivalent in English *run*. For example in the following example *ooTu* is used in the sense of understand: *enakku kaNakku ooTaatu* 'I cannot understand mathematics' (literally means 'For me the mathematics will not run').

7.3. Lexical gaps in the domain of abstracts

7.3.1. Lexical gap in calendar times

Tamil does not have English calendar months such as January, February, March, April, May, June, July, August, September, October, November and December. Tamil only borrowed the names of these months from English.

Tamil society has its own unique system of calendar months. The Tamil new year usually falls in mid-April and the calendar consists of twelve months. Unlike the Gregorian calendar, the number of days in a given month can vary between years. Moreover, Tamil months may even have 32 days. For example, the month of *Vaikasi* had 32 days in 1996 and 31 days in 1998. Similarly, *Aani* had 31 days in 1996 and 32 days in 1998. The following table shows when the first day of each month occurs:

No.	Tamil month	Transliteration	Start date
10.	தை	*tai*	January
11.	மாசி	*maaci*	February
12.	பங்குனி	*pangkuni*	March

1.	சித்திரை	cittirai	April
2.	வைகாசி	*vaikaaci*	May
3.	ஆணி	*aani*	June
4.	ஆடி	*aaTi*	July
5.	ஆவணி	*aavaNi*	August
6.	புரட்டாசி	*uuraTTaaci*	September
7.	ஐப்பசி	*aippaci*	October
8.	கார்த்திகை	*kaartikai*	November
9.	மார்கழி	*maarkazhi*	December

English does not have these indigenous calendar months.

7.3.2. Lexical gap in colour in terms

Another domain of mismatch is colour domain. Colour terms are differentiated by perception of eyes. The mismatch could be due to individual perception and realization. There are basic colour terms Tamil which can be glossed against English. For example, *kaRuppu* 'black', *veLLai* 'white' *civappu* 'red', *niilam* 'blue', *paccai* 'green', *manjcaL* 'yellow'. But the colours such are *aaranj* 'orange', *piravuN* 'brown' *roos* 'rose' are borrowed. Violet is denoted by the colour of a flower, *uutaa* (from *uutaa puu* 'a type of flower'). Tamil colour conscious Tamil speakers coin a lot of new colour terms making use of the colours of the objects known to them; for example *aappiL niRam* 'apple colour'; *ilaip paccai* 'leaf green', etc.

7.4. Lexical gaps in the domain of Relationals

Nida has grouped spatial, temporal, deictic, and logical expressions under relationals. Some of them are realized as prepositions in English where as they are realized as postpositions, case suffixes and particles in Tamil.

8. Consequence of lexical vacuum or gaps

One of the important consequences of lexical gap is borrowing. If the lexical items are not available in a language it resorts to borrowing. Lexical gaps could the reason for code switching too. Non availability of a lexical item denoting a concept makes the language learning of a foreign language difficult. Lexical gaps make the translation between two languages difficult. Lexical gaps make the compilation of a bilingual dictionary difficult. The non-availability of lexical items in a semantic domain makes it difficult to the structure the vocabulary in a neat pattern. The availability of folk taxonomy against a scientific taxonomy distorts the structuring of vocabulary.

9. Conclusion

Lexical vacuum or gap is difficult to define. A language may be self-sufficient with its own vocabulary items. We may feel that certain Tribal languages have minimal number of lexical items to help their speakers conduct their day today life. The people speaking these languages may be self-sufficient in their own territory. But when a language expands its range of use it meets with lexical vacuum or lexical gaps. Language resort to many ways to fill these vacuum or gaps. It is quite natural that languages have culturally motivated lexical items. So a language's culturally bound lexical items may not find its place in another language. As we know Eskimos have several words to denote the concept 'ice'. The people who frequent sea may have a good amount of vocabulary to express the different tides of sea, different types of wind and so on. Lexical gaps are not omnipresent, but not very rare either.

References

Arnold D, Balkan L, Meijer S, Humphreys R.L., Sadler L. 1994. Machine Translation: An Introductory Guide. Blackwell: Blackwell Ltd.

Bentivogli, L And Pianta, E. 2000 Looking for Lexical Gaps. http"//www.itc.it/paper/wordnet-euroalex2000.

Bentivogli, L., Pianta, E. And Pianesi, F. 2000. "Coping with lexical gaps when building aligned multilingual wordnets", in Proceedings of LREC 2000, Athens, Greece.

Chomsky, N. 1965. Aspects of the theory of syntax. Cambridge: The MIT Press.

Chomsky, Noam And Morris Halle. "Some controversial questions in phonological theory." Journal of Linguistics.

Civilikaite, J. 2006. "Lexical gaps: Resolution by functionally complete units of translation." DARBAI IR DIENOS 2006.45.

Conner, D.F.1983 Understanding Semantics. Vol.XI. Exeter, Exete University Press.

FAN, G. L. 1989. Word families, gaps in word families and their uses. Chinese Language Learning, 2, 42-45.

Cruse, D.A. 1986. Lexical Semantics. Cambridge: Cambridge University Press.

Cruse, A. D. 2004. Meaning in language, Oxford: Oxford University Press.

Gazala, H.1995 Translation as Problems and Solutions. Malta: ELGA publication.

Jassen, M. 2004. "Multilingual Lexical Databases, Lexical gaps, and SIMuLLDA." International Journal of Lexicography, Vol. 17, no. 2, 137-154.

Jones, K.S. 1986. Synonymy and Semantic Classification. Edinburgh: Edinburgh University Press.

Kjellmer, G. 2003. "Lexical gaps." Language and computers, vol. 48, 1.149-158.

Löbner, S. 2002 Understanding Semantics. London: Edward Arnold.

Lehrer, A. 1974 Semantic Fields and Lexical Structure. Amsterdam: North-Holland Publishing Company.

Lyons, J.1977 Semantics. Vol.I. Cambridge: Cambridge University Press.

Lyons, J.1981 Language and Linguistics. Cambridge: Cambridge University Press.

Madaiyan, P and Chitraputhran, H. 1986. *mara iaappeyarkaL* (Names of plants). Tamil University, Thanjavur.

Maguire, Susan-Bassnett. 1980 Translation Studies. London: Muethen.

Nida, E.A. 1975a. Language Structure and Translations. Stanford: Stanford University Press.

Nida, E.A. 1975b. Compositional Analysis of Meaning: An Introduction to Semantic Structure. The Hague: Mouton.

Nida, E.A.. 1975.c. Exploring Semantic Structure. The Hague: Mouton

Nida, E.A. 2000 "Principles of Correspondence" In Laurence Venuti (ed.), The Translation Studies Readers. 126-140 (first published in 1964).

Proost, K. 2007. Conceptual structure in lexical items; the lexicalisation of communication concepts in English, German and Dutch. New York: John Benjamins.

Rajendran, S. 1976. Verbs of Cooking in Tamil. Working Papers in Linguistics No.1, Deccan College, Pune, 1976, 8-16.

Rajendran, S. Thampuran R.R. & Sharma, R.C. 1976. A Note on Directional Verbs of Movement. Bulletin of the Deccan College Research Institute (BDCRI) vol 35, 1976, 113-115.
Rajendran, S. 1978. Syntax and Semantics of Tamil Verbs. Ph.D. Thesis. Poona: University of Poona.
Rajendran, S. 1981. Semantic Structure of Tamil Verbs, 13th AIUTTAC *aayvukkoovai* vol. 2, 1981, 305-310
Rajendran, S. 1982. Semantics of Spatial Expressions in Tamil. 14th AIUTTAC *aayvukkoovai* vol. 1982, 262-67.
Rajendran, S. 1982. Verbs of 'Seeing' in Tamil. In BDCRI vol. 41, 1982, 151-159.
Rajendran, S.1982. Semantic structure of Tamil Vocabulary. Post-doctoral Research Report. Poona: Deccan College.
Rajendran, S. 1983. Semantic Structure of Directional Verbs of Movement. Working papers in Linguistics vol 7, Deccan College, Pune, 1983, 19-37.
Rajendran, S. 1983. Coming and Going in Tamil. In: E. Annamalai (edited) To Greater Heights, CIIL, Myosre, 1983, 83-118.
Rajendran, S. 1983. Temporal Expressions in Tamil. BDCRI vol. 42, 1983, 138-147.
Rajendran, S.. 1995. "Towards a Compilation of a Thesaurus for Modern Tamil". South Asian Language Review. 5.1:62-99
Rajendran, S. 1995. Componential Analysis of 'Eating' in Tamil. PILC Journal of Dravidic Studies, vol. 5, no.2, 1995. 175-181.
Rajendran, S. 1995. Toward the Compilation of a Thesaurus for Modern Tamil. South Asian Language Review, vol. 5, no. 1-4, 1995, 62-99.
Rajendran, S. 1996. "The Feasibility of Preparing a Thesaurus using Corpus". Workshop on Indian Language Corpus and its Applications (28, 29 Oct. 1996). Central Institute of Indain Languages, Mysore.
Rajendran, S. 1996-2000. Componential Analysis of Meaning and its Relevance in Finding Translational Equivalents. Tamil Civilization vol. 14-18, March 1996-2000, 95-98.
Rajendran, S. 1996-2000. Componential Analysis of Meaning and its Relevance in Finding Translational Equivalents. Tamil Civilization vol. 14-18, March 1996-2000, 95-98.
Rajendran, S. 2001. *taRkaalattamizhc coRkaLanciyam* [Thesaurus for Modern Tamil]. Thanjavur: Tamil University.
Rajendran, S. 2018. Lexical gap in the vocabulary of Tamil Language. academia. edu. and Research Gate.
Santos D. 1990. Lexical gaps and idioms in Machine Translation. Proceedings of COLING'90, vol 2, 330–335.
Svensen, B. 2009. A handbook of lexicography: The theory and practice of dictionary-making. Cambridge: Cambridge
University Press.
Trask, R. L.1993 A Dictionary of Grammatical Terms in Linguistics. London: Routledge.
Weise, Günter 1988 "Contrastive Linguistics and the Problem of Equivalence in Translation" In Poznň. Vol.XXII PP.187-194.
Wang, X. J. 1989. Empty signs in language. Chinese Learning Monthly, 2, 67-70.
Wang, Quan-Zhi, 2017. Lexical Gaps: Their Filling and Impacts. Journal of Literature and Art Studies, June 2017, Vol. 7, No. 6, 748-754 doi: 10.17265/2159-5836/2017.06.**011**